Complete Mitochondrial Genome of the Small Brown Planthopper, *Laodelphax striatellus* (Delphacidae: Hemiptera), with a Novel Gene Order

Author(s): Nan Song and Ai-Ping Liang
Source: Zoological Science, 26(12):851-860.
Published By: Zoological Society of Japan
DOI: http://dx.doi.org/10.2108/zsj.26.851
URL: http://www.bioone.org/doi/full/10.2108/zsj.26.851

ZOOLOGICAL SCIENCE **26**: 851–860 (2009)

Complete Mitochondrial Genome of the Small Brown Planthopper, *Laodelphax striatellus* (Delphacidae: Hemiptera), with a Novel Gene Order

Nan Song[1,2] and Ai-Ping Liang[1]*

[1]*Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China*
[2]*Graduate School of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing 100039, China*

We determined the first complete mitochondrial genome (mitogenome) sequence from a representative of the insect family Delphacidae, *Laodelphax striatellus*. The 16,513 bp long *L. striatellus* mitogenome encodes 13 putative proteins, two ribosomal RNAs, and 22 transfer RNAs, and contains a putative control region (or A+T-rich region). The nucleotide composition is biased toward adenine and thymine (77.2% A+T), and the amino acid composition is affected to a similar degree by the AT mutational bias. All 13 protein-coding genes (PCGs) start with a typical ATN initiation codon. Eight of 13 PCGs in *L. striatellus* have a complete termination codon (TAA), whereas the remaining five have incomplete termination codons. The anticodons of the *L. striatellus* tRNAs are identical to those in *Drosophila yakuba*, and all tRNAs except for $tRNA^{Ser\text{-}AGN}$ can be folded in the form of a typical cloverleaf structure. The A+T-rich region of *L. striatellus* was found between *srRNA* and $tRNA^{Ile}$, and the entire region was 2040 bp long. The gene content of the *L. striatellus* mitogenome is identical to other completely sequenced insect mitogenomes, while the gene order is different from the common arrangement found in most insects: five tRNA genes and three PCGs in the *L. striatellus* mitogenome have changed positions relative to the ancestral arrangement of mitochondrial genes in *D. yakuba*. Besides describing the above contents, we also aligned the mitogenome sequence of *L. striatellus* with other hemipterans to analyse the phylogenetic relationships of Hemiptera. The results show that Heteroptera is the sister group to all other Hemiptera, and Cicadomorpha is the sister group to the clade Fulgoromorpha+Sternorrhyncha.

Key words: mitochondrial genome, gene order, *Laodelphax striatellus*, Delphacidae, Hemiptera

INTRODUCTION

The typical metazoan mitogenome is a covalently closed circular molecule, ranging in size from 14 to 19 kb (Wolstenholme, 1992; Boore, 1999). The gene content and organization are generally conserved, containing 37 genes: 13 PCGs, 2 ribosomal RNA (rRNA) genes, and 22 transfer RNA (tRNA) genes. In addition, the mitochondrial DNA (mtDNA) molecule usually contains one major non-coding region that is thought to play a role in the initiation of transcription and replication (Wolstenholme, 1992). The arrangements of mitochondrial genes are usually the same or very similar within a phylum but differ substantially among phyla of animals (Boore and Brown, 1998).

Due to its presumed lack of recombination, maternal inheritance, and a relatively rapid mutation rate, complete mtDNA sequences or partial sequences have been extensively used to study population structure and phylogenetic relationships at various taxonomic levels. Furthermore, the arrangement of genes in a mitogenome provides useful data for the study of the evolutionary relationships of insects (Boore et al., 1998; Roehrdanz et al., 2002). Mitochondrial gene order rearrangements appear to be unique, generally rare events that are unlikely to arise independently in separate evolutionary lineages as the result of convergence (Boore, 1999). However, our limited knowledge of the mechanisms responsible for the rearrangement of mtDNA genes limits their broader acceptance for phylogenetic research (Curole and Kocher, 1999).

At present, complete mitogenome sequences have been determined for more than 100 species of insects. Within Hemiptera, a number of complete or nearly complete

* Corresponding author. Phone: +86-10-64807226;
Fax : +86-10-64807099;
E-mail: liangap@ioz.ac.cn
doi:10.2108/zsj.26.851

ABBREVIATIONS

Mitogenome, mitochondrial genome; PCG, protein-coding gene; mtDNA, mitochondrial DNA; *atp6* and *atp8*, ATP synthase subunits 6 and 8; *cytb*, apocytochrome b; *cox1–3*, cytochrome c oxidase subunits 1–3; *nad1–6* and *nad4l*, NADH dehydrogenase subunits 1–6 and 4l; s*rRNA* and *lrRNA*, small and large subunit ribosomal RNA (rRNA) genes; tRNA, transfer RNA; *CR*, putative control region; aa, amino acids; bp, base pairs; ML, maximum likelihood; BI, Bayesian inference.

mitogenome sequences are available in GenBank, from six species of whiteflies (Aleyrodoidea), one psyllid (Psylloidea), two aphids (Aphidoidea), one planthopper (Fulgoroidea), one spittle bug (Cercopoidea), one leafhopper (Membracoidea), and sixteen true bugs (Heteroptera). However, there is no complete mitogenome sequence for delphacid planthoppers (Delphacidae: Hemiptera), which are a diverse group of phytophagous insects with more than 3000 described species worldwide (O'Brien and Wilson, 1985). Among delphacid species, the small brown planthopper, *Laodelphax striatellus* (Delphacidae: Fulgoroidea), is an economically important and widespread insect pest of rice in China, where heavy infestations occur in the middle and lower reaches of the Yangtze River and in the northern part of the country. The adults and nymphs of *L. striatellus* suck rice sap in large numbers. Leaves infested by *L. striatellus* turn yellow, wilt, and even die, resulting in yield loss and a decline in grain quality. In addition, this planthopper also transmits rice viral diseases such as rice black-streaked dwarf virus and rice stripe virus. These rice viral diseases often cause more severe yield losses than the feeding damage. Besides its economic importance, *L. striatellus* is crucial in phylogenetics. The ordinal classification and evolutionary affiliations of higher taxa in Hemiptera have been debated since Linneaus originally established this order in 1758. Fulgoroidea is considered as a pivotal taxon in determining a phylogenetic framework for the Hemiptera. It is thus necessary to determine the complete mitogenome sequence of *L. striatellus* as a representative of Fulgoroidea.

In this study, we sequenced the complete mitogenome of *L. striatellus* and examined mainly its nucleotide composition, codon usage, and genomic arrangement. In addition, the mitogenome sequence of this species analyzed phylogenetically to provide some insight into the relationships among hemipterans.

MATERIALS AND METHODS

Sample and DNA extraction

An adult specimen of *L. striatellus* was collected from Beijing, China. After an examination of external morphology for identification, the specimen was preserved in 100% ethanol and stored at –80°C in the Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences.

Muscle tissue from under the pronotum was homogenized in 2 ml of chilled buffer (220 mM mannitol, 70 mM sucrose, 5 mM Tris, 2 mM EDTA, pH, 8.0), and centrifuged at 800×*g* to pellet the nuclei and cellular debris. After the resultant supernatant was recovered through centrifugation at 3600×*g*, 1 ml of homogenizing mixture was added to the precipitate and centrifuged again to pellet the mitochondria.

The pellet was digested in protease buffer (100 mM Tris, 40 mM NaCl, 2 mM EDTA, 10% SDS, 5 μl of 20 mg/ml proteinase K). The solution was mixed with 250 μl 5.3 M NaCl and centrifuged at 1400×*g*. After 560 μl of isopropanol was added to the supernatant, the mixture was chilled at –20°C for 30 min and pelleted through centrifugation. The pellets were washed with 70% ethanol and stored at –20°C. DNA was dissolved in 100 μl of double-distilled water, and one-tenth dilutions were used as the template in PCR.

PCR amplification, cloning and sequencing

The genome was amplified in overlapping PCR fragments (Fig. 1 shows a schematic map of the amplification fragments, and detailed information on primers is provided in Table 1). Initial rounds of amplification for genomic sequencing were performed by using sets of heterologous primers developed from aligned insect sequences. As more conserved genes were sequenced, primers were chosen based on the sequences obtained. Short fragments were amplified by using Taq DNA polymerase (QIAGEN, China) under the following cycling conditions: 5 min at 94°C, followed by 30 cycles of 50 s at 94°C, 50 s at 50°C, and 1–3 min at 72°C. The final

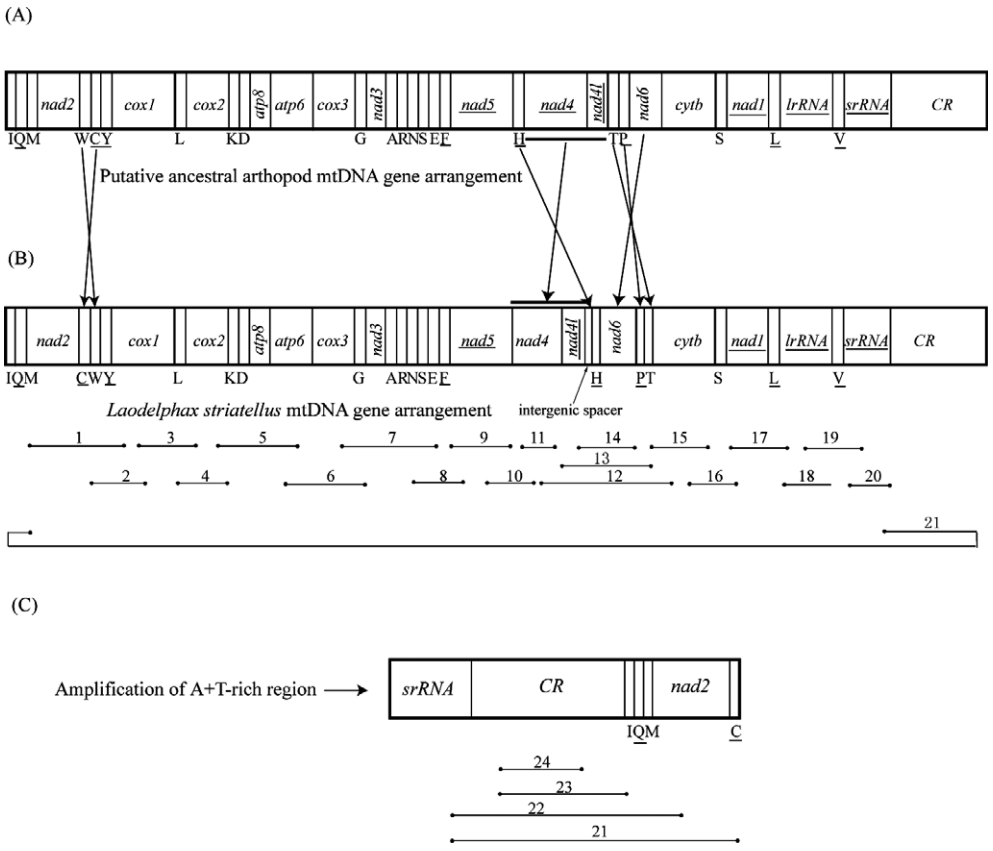

Fig. 1. Linearized representation of the mitogenome showing gene arrangements for (A) the putative ancestral arthropod and (B) *L. striatellus* (Delphacidae: Hemiptera). Arrows indicate the rearrangements of genes or gene clusters. (B, C) Schematic representation of the amplification strategy used for the *L. striatellus* mitochondrial genome. Lines below the linearized genome map represent amplification products.

Table 1. List of primers used in the sequencing of *L. striatellus* mtDNA.

Upstream primers	Sequence (5′→3′)	Downstream primers	Sequence (5′→3′)
F01	AATAAAGCTAATAGGTTCATACCC	R01	TTTATTCGGGGGAATGCTATATC
F02	GTTAAATAAACTAGTAACCTTCAAA	R02	GCTCGTGTATCAACGTCTATACC
F03	AATTGGTGGTTTTGGAAATTG	R03	TATTCATATCTTCAATATCATTGATG
F04	TCTAATATGGCAGATTAGTGCA	R04	ACTATTAGATGGTTTAAGAG
F05	CATTAGATGACTGAAAGCAAGTA	R05	ATGTCCWGCAATTATATTTGC
F06	TTTGCCCATCTWGTWCCTCAAGG	R06	ATTTTTGAAAATCCACATTC
F07	CTATCAGCTTGATACTGACACTTTG	R07	TATTCAGGTAGCCTAATTTTAAAG
F08	AGAGGTATATCACTGTTAATGA	R08	TTAAGGCTTTATTATTTATATGTGC
F09	AGAGGTATATCACTGTTAATGA	R09	TTAGGTTGAGATGGTTTAGG
F10	AAAAGGAAACTGAGCACTTTTAGT	R10	AAATCTTTRATTGCTTATTCTTC
F11	CCAGAAGAACATAANCCATG	R11	TGAGGTTATCARCCTGAACG
F12	ATAGTAGGTCCTTCTACATGAGC	R12	TTTCTACGGAGAATCCTCCTCA
F13	AATAATGTTTACGAACCCAA	R13	CCTAGGATTGAACCAAAATTTCA
F14	CGTTCAGGCTGATAACCTCA	R14	TAGTTTTGGATATTAAAGATGC
F15	CTCATACTGATGAAATTTTGGTTC	R15	TTCTACTGGTCGTGCTCCAATTCA
F16	TCCATATTCAACCAGAATGATA	R16	TTTGTTTCCTGGTCTTGGG
F17	AGGAAAGGAACCACGAACCCA	R17	ATACCTTAGGGATAACAGCGTGA
F18	CCGGTCTGAACTCAGATCAT	R18	ATTTATTGTACCTTTTGTATCAG
F19	CCTTTGTACAGTTAAAATACTGC	R19	AATTATGTACATATCGCCCTTC
F20	GTAAAYCTACTTTGTTACGACTT	R20	GTGCCAGCAAYCGCGGTTATAC
F21	ATAATAGGGTATCTAATCCTAGT	R21	ARCTTTGAAGGYTAWTAGTTT
F22	identical to F21	R22	CCTGTTGATTAAAACCATTGGG
F23	AATCTAAAATAATTTATAATATATACTA	R23	GGATCCAAATCCCCCTTTTA
F24	identical to F23	R24	TCGTGTGCAAGGAATTTGG

F, forward; R, reverse.

elongation step was continued for 10 min at 72°C. For large fragments, long PCRs were performed using Long Taq DNA polymerase (QIAGEN, China) with the following cycling conditions: 2 min at 96°C, followed by 30 cycles of 10 s at 98°C, and 10 min at 68°C. The final elongation was continued for 10 min at 72°C. These PCR products were analyzed by 1.0% agarose gel electrophoresis.

PCR products ~1200 bp long (fragments 1–11, 13–20, and 22–24 in Fig. 1) were directly sequenced after purification, but PCR products 1.2–3.2 kb long (fragments 12 and 21 in Fig. 1) were cloned into pBS-T Easy vector (QIAGEN, China) and the resultant plasmid DNA was isolated by using the TIANprp Midi Plasmid Kit Purification System (QIAGEN, China). For each larger PCR product, at least two independent clones were sequenced to ensure that we obtained the true sequence. Internal primers, necessary to complete the sequencing of cloned fragments, were designed by using each amplified fragment as a template. DNA sequencing was performed with a BigDye Terminator Cycle Sequencing Kit and an ABI 3730XL Genetic Analyzer (PE Applied Biosystems, USA). All fragments were sequenced from both strands.

Sequence assembly, annotation, and analysis

Raw sequence files were proofread and aligned into contigs in BioEdit version 7.0.5.3 (Hall, 1999). Contig sequences were checked for ambiguous base calls, and only non-ambiguous regions were used for annotation. Sequences alignment, genome assembly, and nucleotide composition calculations were all conducted with Mega 4 (Tamura et al., 2007). The locations of PCGs and rRNA genes were identified by determining sequence similarity with homologs in other insects, while tRNA genes were identified by using the tRNAscan-SE server (Lowe and Eddy, 1997). The tRNAs not found by tRNAscan-SE were identified through comparing them to the regions coding these tRNAs in other insects. Potential secondary structural folds in the A+T-rich region of the genome were predicted by using Mfold v. 3.2 (http://www.bioinfo.rpi.edu/ applications/ mfold/) (Zuker, 2003).

Phylogenetic analysis

A phylogenetic analysis was carried out based on 29 complete or nearly complete mitogenome nucleotide sequences from hemipteran insect species, including *L. striatellus* and *Geisha distinctissima* (Song and Liang, 2009). Alignments of 13 PCGs based on amino acid sequences were made and concatenated in Mega4, with the stop codons of the PCGs excluded. This data set was used to reconstruct phylogenetic trees by maximum likelihood (ML) and Bayesian analyses, respectively. ML analyses were conducted in PAUP v. 4b10 (Swofford, 2002). The GTR+I+G model was selected as the best-fitting model for the nucleotide sequences, and the parameters estimated by Modeltest 3.7 (Posada and Crandall, 1998) were as follows: base frequencies of A=0.2646, C=0.2210, G=0.2211, T=0.2933; invariable sites of I=0.2298; and a gamma distribution shape parameter of variable sites of G=0.5140 (the above parameters were for the all-sites data set). The number of bootstrap replicates was 100. Bayesian analyses were conducted with MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001) with the following options: four independent Markov chains, three million generations, tree sampling every 100 generations, and the first 25% discarded as burn-in. Another data set including only the 1st and 2nd codon positions of PCGs was also used to reconstruct the phylogeny by the same methods.

RESULTS

Genome composition

The complete mitogenome of *L. striatellus* is circular and is 16,513 bp long. The sequence data have been deposited in the GenBank database under accession number FJ360695. The sequence analysis revealed a gene content typically found in metazoan mitogenomes: 13 PCGs, 22 tRNA genes, the large and small rRNA (*lrRNA* and *srRNA*) subunits, and the A+T-rich region (Fig. 1; Table 2).

The whole genomic organization of *L. striatellus* is very

compact, with only 131 nucleotides dispersed in 10 intergenic spacers ranging in size from 1 to 72 bp. The two largest intergenic spacers appear to be of significant length: 34 bp between $tRNA^{Cys}$ and $tRNA^{Trp}$, and 72 bp between *nad4l* and $tRNA^{His}$. The contiguous genes in *L. striatellus* overlap at 13 boundaries by a total of 98 bases, with the two largest measuring 20 bp located between *cox2* and $tRNA^{Lys}$, and between $tRNA^{Ser\text{-}UCN}$ and *nad1*.

The nucleotide composition of the *L. striatellus* mitogenome is heavily biased toward adenine and thymine (77.2%), as with other insect mtDNA sequences. Nucleotide compositions for features within the genome are listed in Table 3. The average A+T content of tRNA (78.1%) and rRNA genes (77.0%) is larger than that of protein genes (76.0%), but the bias in the third codon positions (majority strand, 82.4%; minority strand, 88.7%) of the protein genes is as extreme as that in the A+T-rich region (83.1%). Interestingly, the second codon positions in the majority strand harbor a higher ratio of T (44.4%) than A (23.5%), whereas

Table 2. Summary of the mitochondrial genes in *L. striatellus*.

Gene	Direction	Span/bp	Size/bp	Anticodon	Start	Stop
$tRNA^{Ile}$	F	1–66	66	GAT(30–32)		
$tRNA^{Gln}$	R	69–134	66	TTG(104–106)		
$tRNA^{Met}$	F	134–198	65	CAT(167–169)		
nad2	F	199–1176	978		ATT	TAA
$tRNA^{Cys}$	R	1180–1241	62	GCA(1210–1212)		
$tRNA^{Trp}$	F	1276–1343	68	TCA(1306–1308)		
$tRNA^{Tyr}$	R	1347–1407	61	GTA(1376–1378)		
cox1	F	1412–2945	1534		ATG	T-tRNA
$tRNA^{Leu\text{-}UUR}$	F	2946–3011	66	TAA(2975–2977)		
cox2	F	3012–3695	684		ATT	TAA
$tRNA^{Lys}$	F	3676–3746	71	CTT(3706–3708)		
$tRNA^{Asp}$	F	3747–3808	62	GTC(3777–3779)		
atp8	F	3809–3970	162		ATT	TAA
atp6	F	3964–4618	655		ATG	T-cox3
cox3	F	4619–5339	781		ATG	T-tRNA
$tRNA^{Gly}$	F	5400–5460	61	TCC(5430–5432)		
nad3	F	5461–5811	351		ATT	TAA
$tRNA^{Ala}$	R	5811–5871	61	TGC(5840–5842)		
$tRNA^{Arg}$	F	5876–5939	64	TCG(5904–5906)		
$tRNA^{Asn}$	F	5939–6002	64	GTT(5969–5971)		
$tRNA^{Ser\text{-}AGN}$	F	5999–6062	64	GCT(6022–6024)		
$tRNA^{Glu}$	F	6057–6118	62	TTC(6087–6089)		
$tRNA^{Phe}$	R	6118–6182	65	GAA(6148–6150)		
nad5	R	6183–7914	1732		ATT	T-tRNA
nad4	R	7901–9238	1338		ATG	TAA
nad4l	R	9232–9504	273		ATG	TAA
$tRNA^{His}$	R	9577–9637	61	GTG(9608–9610)		
nad6	F	9637–10147	511		ATA	T-tRNA
$tRNA^{Pro}$	R	10148–10209	62	TGG(10177–10179)		
$tRNA^{Thr}$	F	10214–10276	63	TGT(10244–10246)		
cytb	F	10281–11384	1104		ATG	TAA
$tRNA^{Ser\text{-}UCN}$	F	11370–11440	71	TGA(11403–11405)		
nad1	R	11421–12374	954		ATG	TAA
$tRNA^{Leu\text{-}CUN}$	R	12376–12439	64	TAG(12408–12410)		
lrRNA	R	12440–13658	1219			
$tRNA^{Val}$	R	13659–13726	68	TAC(13687–13689)		
srRNA	R	13727–14473	747			
A+T-rich region		14474–16513	2040			
repeat region		15593–16299	707			

Table 3. Nucleotide composition for features of the mitogenome of *L. striatellus*.

Genome feature	No. of nuceotides	A (%)	T (%)	G (%)	C (%)	A+T (%)
Whole genome (Majority strand)	16513	43.2	34.0	9.3	13.5	77.2
Majority strand protein genes	6741	37.2	36.7	10.0	16.2	73.9
1st codon position	2247	41.2	30.2	14.8	13.8	71.4
2nd codon position	2247	23.5	44.4	11.9	20.2	67.9
3rd codon position	2247	47.0	35.4	3.1	4.5	82.4
Minority strand protein genes	4287	25.9	53.3	12.6	8.1	79.2
1st codon position	1429	27.6	48.8	15.8	7.8	76.4
2nd codon position	1429	19.6	53.1	14.3	12.9	72.7
3rd codon position	1429	30.5	58.2	7.8	3.6	88.7
Whole tRNA (Majority strand)	1416	42.8	35.3	10.2	11.7	78.1
Majority strand tRNA	907	43.3	36.4	10.6	9.7	79.7
Minority strand tRNA	509	33.4	41.8	15.3	9.4	75.2
rRNAs (Minority strand)	1966	34.5	42.5	15.1	7.8	77.0
tRNAs (Majority strand)	1416	42.8	35.5	10.2	11.7	78.3
A+T-rich region (Majority strand)	2040	42.5	40.6	10.0	6.1	83.1
repeat region	707	38.5	32.5	19.2	9.8	71.0

Table 4. Relative synonymous codon usage in *L. striatellus*. RSCU values were calculated for all 13 protein coding genes, and then separately for genes encoded only on the major strand and only on the minor strand. Values in bold font represent the most commonly used codon for the given amino acid. This analysis excluded termination codons.

a.a.	Codons	All	Majority	Minority	a.a.	Codons	All	Majority	Minority
K	AAA	**1.75**	**1.85**	**1.38**	L	UUA	**3.97**	**3.65**	**4.33**
	AAG	0.25	0.15	0.62		UUG	0.46	0.14	0.82
N	AAU	**1.55**	**1.40**	**1.82**	P	CCU	**1.76**	1.38	**3.03**
	AAC	0.45	0.60	0.18		CCC	0.90	1.13	0.14
Q	CAA	**1.77**	**1.84**	**1.60**		CCA	1.31	**1.50**	0.69
	CAG	0.23	0.16	0.40		CCG	0.03	0.00	0.14
H	CAU	**1.13**	**0.94**	**1.85**	A	GCU	**2.05**	**1.47**	**3.52**
	CAC	0.87	1.06	0.15		GCC	0.77	1.02	0.16
E	GAA	**1.77**	**1.89**	**1.50**		GCA	1.14	1.46	0.32
	GAG	0.23	0.11	0.50		GCG	0.05	0.06	0.00
D	GAU	**1.45**	**1.16**	**1.81**	S	UCU	2.07	1.46	**2.98**
	GAC	0.55	0.84	0.19		UCC	0.61	0.85	0.26
Y	UAU	**1.48**	**1.03**	**1.84**		UCA	**2.54**	**3.27**	1.46
	UAC	0.52	0.97	0.16		UCG	0.13	0.04	0.26
W	UGA	**1.80**	**1.91**	**1.45**	S	AGU	0.99	0.50	**1.73**
	UGG	0.20	0.09	0.55		AGC	0.06	0.04	0.10
C	UGU	**1.87**	**1.54**	**2.00**		AGA	**1.52**	**1.81**	1.10
	UGC	0.13	0.46	0.00		AGG	0.06	0.04	0.10
M	AUA	**1.74**	**1.84**	**1.56**	R	CGU	**2.00**	1.27	**2.89**
	AUG	0.26	0.16	0.44		CGC	0.00	0.00	0.00
I	AUU	**1.68**	**1.57**	**1.94**		CGA	1.60	**2.55**	0.44
	AUC	0.32	0.43	0.06		CGG	0.40	0.18	0.67
F	UUU	**1.71**	**1.48**	**1.90**	G	GGU	**1.76**	1.08	**2.65**
	UUC	0.29	0.52	0.10		GGC	0.14	0.08	0.22
L	CUU	**0.70**	0.71	**0.68**		GGA	1.55	**2.33**	0.54
	CUC	0.13	0.24	0.00		GGG	0.54	0.50	0.59
	CUA	0.68	**1.14**	0.16	V	GUU	**1.86**	1.53	**2.22**
	CUG	0.06	0.12	0.00		GUC	0.05	0.05	0.06
T	ACU	1.66	1.43	**2.56**		GUA	1.80	**2.22**	1.33
	ACC	0.55	0.55	0.56		GUG	0.29	0.20	0.39
	ACA	**1.69**	**1.90**	0.89	TER	UAA	–	–	–
	ACG	0.09	0.12	0.00		UAG	–	–	–

Note: RSCU values were calculated for all 13 genes and then major strand encoded only and minor strand encoded only genes. Values in bold type represent the most commonly used codon for the given amino acid. This analysis excludes the termination codons.

Table 5. Sequences of *L. striatellus* tRNAs, with landmarks in the secondary structure. Underlined nucleotides form base pairs. Anticodons are indicated in bold font.

tRNA	Acceptor arm		DHU arm	Anticodon arm	Variable loop	TψC arm	Acceptor arm
tRNA^{Ile}	[AGAAAGA]	UG	(CCUGAUUAAAGG)	A(GUAUUCU**GAU**AGAAUAA)	AUCAA	(GAAAUAAAUUUUUC)	[UCUUUCUA]
tRNA^{Gln}	[UAUAAAA]	UGU	(UGUUUAGCA)	UA(AAGAAUU**UUG**AUUUCUU)	AGGU	(AUUAGUUUAUUCUAAU)	[UUUUAUAA]
tRNA^{Met}	[AAAAAGA]	UA	(AGCUAAAAUUUAAGCU)	A(UUGGGCC**CAU**AACCCAA)	CUAU	(GAUUAUUAUC)	[UCUUUUUA]
tRNA^{Cys}	[UGACUUA]	UAU	(UCAAAAAUGA)	UU(AUAAAUU**GCA**AAUUUAU)	AGGU	(GAAUUUAAUUC)	[UAGGUCUU]
tRNA^{Trp}	[AAGGAUU]	UA	(AGUUAAAUAAACU)	A(GUAACCU**UCA**AAGUUAA)	AAAU	(AGAUUAAAAAAAAUUU)	[AAACCUUA]
tRNA^{Tyr}	[AAUAGGG]	UGU	(CUGAUUUAGG)	UG(AUAAACU**GUA**AAUUUAU)	UUAA	(GGGUCCUCCC)	[UCCUAUUA]
tRNA$^{Leu\text{-}UUR}$	[UCUAAAA]	UG	(GCAGAAUAGUGU)	A(AUGAAUU**UAA**AAUUCAU)	UUAU	(GAAAUCCUUUAUUUC)	[UUUUAGAA]
tRNA^{Lys}	[CAUUAAG]	UG	(ACUGAAAUUAAGU)	A(AUGGUCU**CUU**AAACCAA)	UUUAU	(AGCAAGUUAAAGAAUGCU)	[CUUAAUGA]
tRNA^{Asp}	[GAGAAAU]	UA	(GUUUAAAAAAAAC)	A(UUAACAU**GUC**AAAUUAA)	AAUU	(ACUUGAAAGU)	[AUUUCUUU]
tRNA^{Gly}	[UUUUCUU]	UA	(GUAUAAAAAGUAU)	A(UUUAACU**UCC**AAUUAAA)	AGGU	(UUUAAUUAA)	[AAGAAAAA]
tRNA^{Ala}	[AGGAGAA]	UA	(GUUAACUAUAAC)	A(UUUAAAU**UGC**AAUUAAA)	AAGU	(ACAUAUUUGU)	[UUCUCUUA]
tRNA^{Arg}	[AGUAAAG]	AA	(GUAAAAUAUAC)	A(AUUAAUU**UCG**ACUUAAU)	UUAA	(GAGAUUAAAAUCUC)	[CUAUACUU]
tRNA^{Asn}	[UUAACUA]	AA	(GCUAAAAAGAAGC)	A(UUUUACU**GUU**AAUAAAA)	AAAAU	(GAUUAAAAAUC)	[UAGUUAAA]
tRNA$^{Ser\text{-}AGN}$	[UAAACGA]		(AGUUUAAAA)	(GGAAGCC**GCU**AACUUCC)	UAAA	(AACUUUAAAUUGUUAAGUU)	[UUGUUUAU]
tRNA^{Glu}	[GUUUAUG]	UA	(GUUUAAAUAAAAC)	C(AGUUAUU**UUC**AAUAACU)	AAAU	(AAUUAUAUUU)	[CUUAAACU]
tRNA^{Phe}	[AUUCAGG]	UA	(GCCUAAUUUUAAAGU)	U(GAUCAUU**GAA**AAUGCUU)	ACAA	(ACAGUUUAUGU)	[UUUGAAUA]
tRNA^{His}	[UCAUAUG]	UA	(GUUUCUUAAC)	A(UAAUAGG**GUG**UUUAUUA)	CAUA	(CAAUUAGUAUUG)	[CAUUUGAU]
tRNA^{Pro}	[CAGUUAA]	UA	(GUUUAUUAAAAAU)	U(UUAGUUU**UGG**AUAUUAA)	AGAU	(GCAUGUGUGU)	[UUGACUGA]
tRNA^{Thr}	[AAUUUAA]	UA	(GUUUAUUAAAAAC)	A(UCGAUUU**UGU**AAAUCGA)	AGUU	(AGAAAAAUUCU)	[UUAAAUUA]
tRNA$^{Ser\text{-}UCN}$	[AAAAUUA]	CG	(AAUUAGUUAAAUUAUU)	G(UGUAUCU**UGA**AAAUACA)	ACU	(UAAAAUUAAUUUUUUAU)	[UAAUUUUA]
tRNA$^{Leu\text{-}CUN}$	[GCCAAUU]	UG	(GCAGAUAAGUGU)	U(UUAAAUU**UAG**AAUUUAA)	AAAU	(GUAUAACAUAUAC)	[AUUUGGUA]
tRNA^{Val}	[CAGAUAG]	A	(UUCAUGGCGAA)	AU(AAUUUUU**UAC**AAUAAUU)	ACUUU	(AACUGUUCAAUUCAGUU)	[UUGUCUGA]

Note: Underlined nucleotides form base pairs. Anticodons are represented in bold type.

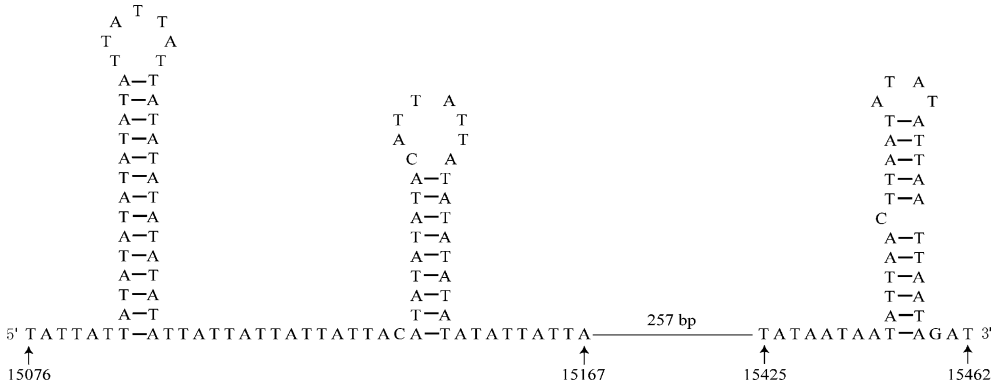

Fig. 2. Potential hairpin structures in the *L. striatellus* A+T-rich region. The numbers represent locations in the whole mitogenome.

the first and third codon positions show the opposite bias. We also noted that G is underrepresented compared to C in the second codon positions of majority strand protein genes, while the first and third codon positions of the minority strand are biased towards higher G content.

Protein-coding genes and codon usage

The mitogenome of *L. striatellus* encodes the regular set of 13 PCGs found with few exceptions in all animal mitogenomes: the majority strand encodes nine PCGs (*nad2*, *nad3*, *nad6*, *cox1*, *cox2*, *cox3*, *atp6*, *atp8* and *cytb*), and the minority strand encodes four PCGs (*nad1*, *nad4*, *nad4l* and *nad5*). Translation initiation and termination codons of the 13 PCGs in *L. striatellus* are summarized in Table 2. All PCGs were observed to have a putative, in-frame ATR (methionine) or ATT (isoleucine) codon as a start signal. Seven start codons are coded by ATG (*cox1*, *atp6*, *cox3*, *nad4*, *nad4l*, *cytb* and *nad1*), five by ATT (*nad2*, *cox2*, *atp8*, *nad3* and *nad5*), and one by ATA (*nad6*). Eight of 13 PCGs in *L. striatellus* have a complete TAA termination codon. The remaining genes have incomplete termination codons (T in *cox1*, *atp6*, *cox3*, *nad5*, and *nad6*). These incomplete termination codons are thought to be extended to TAA during the maturation of transcript, a phenomenon commonly observed in metazoan mitochondrial genes (Bae et al., 2004; Junqueira et al., 2004; Clary and Wolstenholme, 1985).

The relative synonymous codon usage (RSCU) values for *L. striatellus* are shown in Table 4. Leucine, phenylalanine, isoleucine, and methionine are the four most frequently used amino acids, respectively accounting for 13.11%, 11.24%, 10.64%, and 6.56% in the *L. striatellus* mitochondrial proteins. The four most frequently used codons, TTA (leucine), TTT (phenylalanine), ATT (isoleucine), and ATA (methionine), are all composed wholly of A and/or T. This demonstrates that the amino acid composition is affected to a similar degree by the AT mutational bias.

Transfer RNA and ribosomal RNA genes

In the mitogenome of *L. striatellus*, 19 of the expected

22 tRNA genes were identified by tRNAscan-SE Search Server version 1.21 (Lowe and Eddy, 1997). The structures of the other three tRNAs ($tRNA^{Ser\text{-}AGN}$, $tRNA^{His}$, and $tRNA^{Ser\text{-}UCN}$) were determined by visual inspection, and by alignment with other insect tRNA genes. All tRNAs have the typical cloverleaf structure except for $tRNA^{Ser\text{-}AGN}$ (Table 5). In this case, the dihydrouridine (DHU) arm cannot form, as in several other metazoan species, including some insects (Wolstenholme, 1992). The tRNA genes in *L. striatellus* range in size from 61–71 bp. The anticodons of the *L. striatellus* tRNAs are identical to those in *D. yakuba* (Clary and Wolstenholme, 1985), *Triatoma dimidiata* (Dotson and Beard, 2001), and *Philaenus spumariu* (Stewart and Beckenbach, 2005).

The two rRNA genes (*lrRNA* and *srRNA*) in the *L. striatellus* mitogenome are located between $tRNA^{Leu\text{-}CUN}$ and $tRNA^{Val}$, and between $tRNA^{Val}$ and the A+T-rich region, respectively (Fig. 1; Table 2). The lengths of the *lrRNA* and *srRNA* genes are respectively 1218 bp and 747 bp, similar to those in *P. spumarius* (1245 bp *lrRNA* and 754 bp *srRNA*). The A+T contents of the *lrRNA* and *srRNA* genes are respectively 78.5% and 74.6% in *L. striatellus*.

A+T-rich region

The A+T-rich region initiates replication in both vertebrates and invertebrates, and the reduced G+C content is one of the most outstanding features of this region (Boore, 1999). The A+T-rich region of *L. striatellus* was found between *srRNA* and $tRNA^{Ile}$, and the entire reigon was 2040 bp long, with an A+T content of 83.1%. The length of the *L. striatellus* A+T-rich region is the second reported for any hemipteran species, after *Trialeurodes vaporariorum* (3729 bp) (Thao et al., 2004). The A+T content (83.1%) in the *L. striatellus* A+T-rich region is higher than in *T. dimidiata* (66%) (Dotson and Beard, 2001) and *P. spumariu* (78.9%) (Stewart and Beckenbach, 2005).

Gene order

The mitogenome of *D. yakuba* was the first insect mitogenome sequence obtained. The mitochondrial gene order originally observed in *D. yakuba* has been demonstrated to be the ancestral gene order in hexapods and crustaceans, as representatives with this gene arrangement are observed in each of these groups (Clary and Wolstenholme, 1985; Hwang et al., 2001; Nardi et al., 2003b). A novel aspect of our study is the finding that there have some gene rearrangements in the mitogenome of *L. striatellus*. Three PCGs (*nad4*, *nad4l* and *nad6*) and five tRNA genes ($tRNA^{Cys}$-$tRNA^{Trp}$, $tRNA^{His}$ and $tRNA^{Pro}$-$tRNA^{Thr}$) are translocated or inverted compared with the putative ancestral arthropod gene arrangement demonstrated by *D. yakuba* (Fig. 1). However, we did not find the same gene rearrangements in *G. distinctissima*, which is a close relative of *L. striatellus*; *G. distinctissima* has the same gene order as *D. yakuba*.

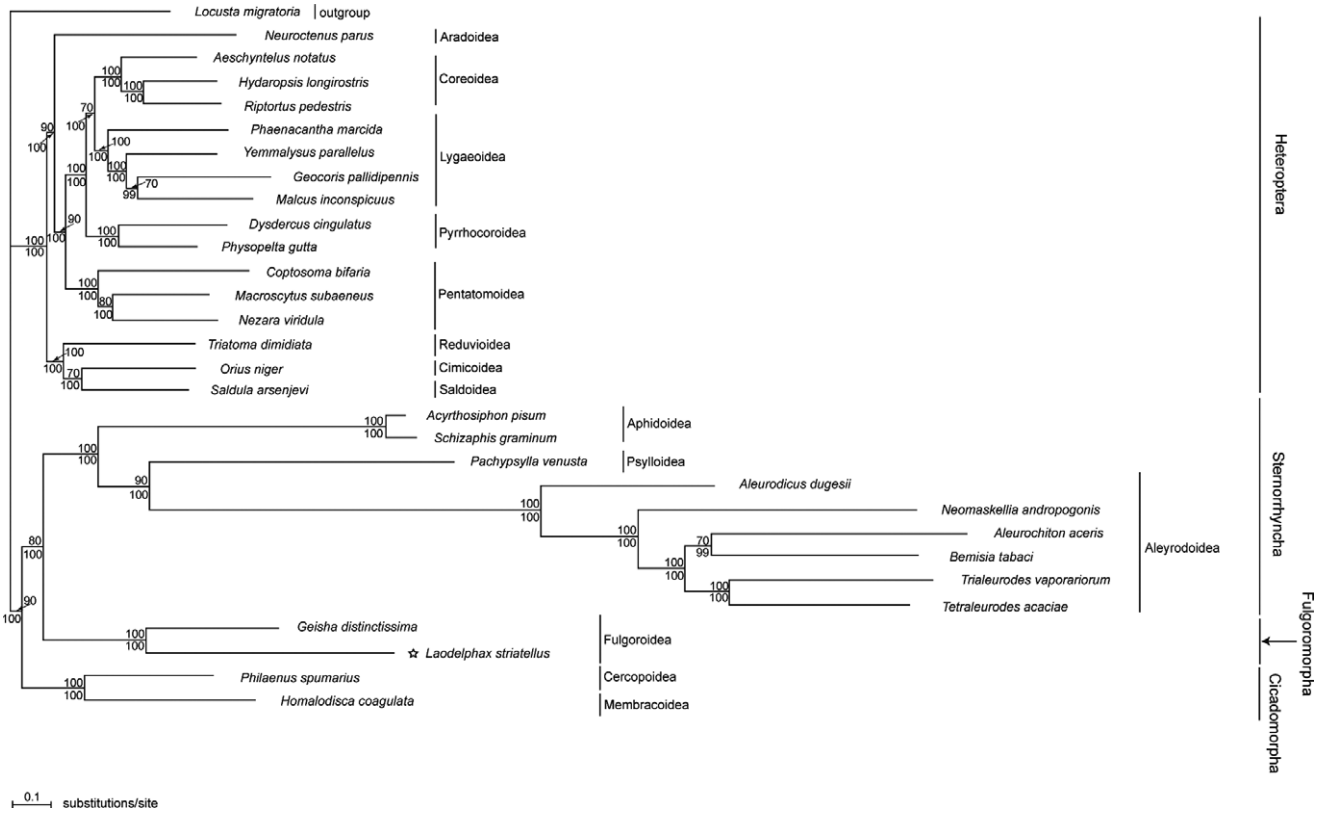

Fig. 3. Phylogenetic tree for Hemiptera resulting from ML and BI analyses of the all-sites data set (including all three codon positions). This tree is identical in topology to trees resulting from analyses of the partial-sites data set, which included only 1st and 2nd codon positions. Values above branches are bootstrap values from the maximum likelihood analysis; values below branches are posterior probabilities from the Bayesian analysis. ☆, the species for which a mitogenome with gene rearrangements was determined in this study.

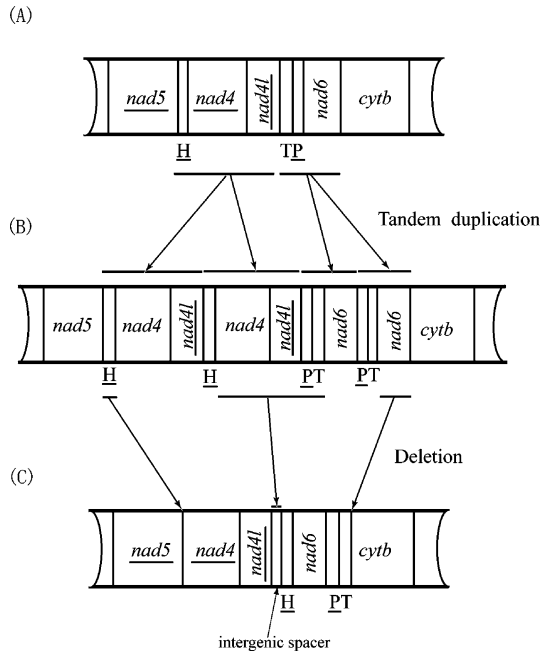

Fig. 4. Proposed duplication-and-deletion mechanism for *L. striatellus*. (A) The typical mtDNA gene order in arthropods. (B) Tandem duplication in the mitogenome. (C) The gene order in *L. striatellus*. Tandem duplication presumably occurred in the *tRNA-^{His}*[H]–*nad4*–*nad4l*–*tRNA^{Thr}*[T]–*Trna^{Pro}*[P]–*nad6* region (the typical mtDNA gene order in arthropods), followed by deletions of redundant genes (*tRNA^{His}*[H]; *nad4*–*nad4l*–*tRNA^{Thr}*[T]–*Trna^{Pro}*[P]; *nad6*). The 72-bp intergenic spacer sequence may have resulted from incomplete deletion events.

Phylogenetic analyses

The all-sites data set contained 11,499 nucleotides sites, and the partial-sites data set including only the 1st and 2nd codon positions contained 7666 nucleotides sites for each of the 30 taxa included. All four phylogenetic analyses conducted with ML and Bayesian inference (BI) generated the same topology (Fig. 3). *Laodelphax striatellus* grouped with *G. distinctissima*. Also, the aphid *Acyrthosiphon pisum* grouped with the aphid *Schizaphis graminum*, and the spittlebug *P. spumarius* with the sharpshooter *Homalodisca coagulata*. Significant nucleotide divergences between species in Aleyrodoidea and other hemipterans resulted in a long branch for Aleyrodoidea in the BI tree. Nonetheless, a monophyletic Sternorrhyncha, (Aphidoidea+(Psylloidea+Aleyrodoidea)), was recovered, and the clade Sternorrhyncha+Fulgoromorpha was strongly supported. In addition, a monophyletic Pentatomomorpha, (Aradoidea+(Pentatomoidea+(Pyrrhocoroidea+(Lygaeoidea+Coreoidea)))), was recovered in both the ML and BI analyses. The inferred genealogical proximities of hemipteran lineages supported the hypothesis of (Heteroptera+(Cicadomorpha+(Fulgoromorpha+Sternorrhyncha))).

DISCUSSION

Mitogenomes are very important to several scientific disciplines, such as animal health, molecular evolution, phylogenetics, and population genetics. To date, the complete or nearly complete mitogenome has been sequenced from more than 100 insect species, including 27 species in the order Hemiptera. None, however, was available from the planthopper family, Delphacidae, species of which are important economically and phylogenetically.

Genome organization and base composition

The 16,513 bp long *L. striatellus* mitogenome is the third largest complete mitogenome sequence reported to date in Hemiptera, behind the kissing bug (*T. dimidiata*) and the white fly (*T. vaporariorum*) (Dotson and Beard, 2001; Thao et al., 2004). The gene content is identical to that in other completely sequenced insect mitogenomes, while the gene order differs from the common arrangement found in most insects: five tRNA genes and three PCGs in this genome have changed positions relative to the ancestral arrangement of mitochondrial genes in *D. yakuba*.

Similarly to *L. striatellus*, other insect species also have intergenic spacer sequences in the mitogenome. For example, *Pyrocoelia rufa* contains an intergenic spacer sequence 1724 bp long, composed of twelve 134-bp tandem repeats plus one incomplete 116-bp repeat (Bae et al., 2004). Also, *T. dimidiata* contains an intergenic spacer sequence 314 bp long, which possibly encodes an unknown gene, because the complementary strand has a complete start methionine and stop codon (Dotson and Beard, 2001). Excluding these examples, the length of intergenic spacers reported in other insects is usually less than 50 bp. In most cases, intergenic spacer sequences consist of only 1 or 2 bp. The total sizes of overlapping regions in the contiguous genes in other insects range from 20 bp (*Bombyx mori*) to 152 bp (*Anopheles quadrimaculatus*) (Yukuhiro et al., 2002).

Considering separately the two strands in the mitogenome of *L. striatellus*, an asymmetrical compositional bias can be observed, and is most evident in comparing gene sequences on the opposite strands: genes encoded on the majority strand (*nad2*, *nad3*, *nad6*, *cox1*, *cox2*, *cox3*, *cytb*, *atp6* and *atp8*) have a comparable AT content (37.2% A; 36.7% T, on average), while genes on the minority strand (*nad1*, *nad4*, *nad5*, *nad4l*, *srRNA* and *lrRNA*) display a strong bias towards higher T than A (25.9% A; 53.3% T). This suggests that the A bias in the overall genome is driven by mutational pressure on the minority strand, which favors T over A. This feature is similar in other insects, for instance in *Reticulitermes* (Cameron and Whiting, 2007).

The start codon for *cox1* is highly variable across insects, and is frequently noncanonical, coding for amino acids other than methionine (Bae et al., 2004; Junqueira et al., 2004; Kim et al., 2006). The *L. striatellus cox1* sequence does not share this feature, but uses the conventional methionine as the first amino acid (start codon ATG). Similarly, other insects (e.g., *P. spumarius*, *T. dimidiata*, and *Pachypsylla venusta*) also use typical methionine (ATG) as the first amino acid for *cox1* (Dotson and Beard, 2001; Thao et al., 2004; Stewart and Beckenbach, 2005). This seems to suggest that the abnormalities are taxon specific.

Two reading frame overlap in the protein-coding strand for *L. striatellus*: a 7-bp overlap (ATGTTAA) between the termination of *atp8* and the initiation of *atp6* on the majority strand, and a 7-bp (ATGTATA) overlap between *nad4l* and *nad4* on the minority strand. Overlaps at these gene boundaries and of this length are quite common among other

insects (Dotson and Beard, 2001; Stewart and Beckenbach, 2005; Krzywinsk et al., 2006).

A feature of most arthropod genomes sequenced to date is that the bias toward the nucleotides A and T also leads to a bias in the amino acids used. This is clearly reflected in the proportions of amino acids with A or T versus C or G in the third codon position (Table 3). *Laodelphax striatellus* appears to share the bias of all arthropod genomes toward amino acids encoded by AT-rich codons. The twofold degenerate amino acids demonstrate a clear bias towards codons with A or T in the third codon position for both strands. For fourfold degenerate amino acids, *L. sriatellus* show a bias toward A in the third codon position for PCGs encoded on the majority strand, and a bias toward T in the third codon position for PCGs encoded on the minority strand. This reflects mutation patterns in the mitogenome, as nucleotides in the third codon position are under the least selective pressure.

A+T-rich region

In the A+T-rich region of *L. striatellus*, a 469-bp region bordered by the small ribosomal unit is heavily biased toward A+T (89.5%) and contains two 8- or 9-bp poly-thymine stretches at the 5' end. Immediately following this region is a 21-bp poly-thymine stretch. This poly-thymine stretch is relatively well conserved across insects (Zhang and Hewitt, 1997). In *Anabrus simplex*, this stretch spans approximately 20 thymine nucleotides, and in *Locusta migratoria* it is roughly 16 bp long and includes two guanines and two adenines. Researchers have speculated that this poly-thymine stretch is involved in transcriptional control or is a site for initiation of replication (Lewis et al., 1994; Zhang et al., 1995). Following the 21-bp poly-thymine stretch is a 626-bp [TA(A)]n-like sequence. In this region, several DNA segments have the potential to form stem-and-loop structures. These structures are formed by stems with perfect matches of varying numbers of nucleotide pairs and loops of various sizes (Fig. 2). These stem-and-loop structures in the A+T-rich region have also been detected in other insects (Zhang et al., 1995; Schultheis et al., 2002), and their presence in the conserved sequences of diverse insects suggests they are an origin of replication for the secondary strand (Zhang et al., 1995).

Varying copy numbers of tandemly repeated elements was reported as characteristic of insect A+T-rich regions (Zhang and Hewitt, 1997). The A+T-rich region of *L. striatellus* harbors a 21-bp repeat unit (5'- TGTCACGATTTTTG-GAAAAAA-3') on the majority strand, repeated 34 times. Repeat unit 3, 11, 19, 27, and 32 have a deletion of T at position 13. In addition, repeat unit 1 has a deletion of T at position 3 and an insertion of CA at position 5, and repeat unit 34 has transversions of T to G, G to A, and A to G at positions 13, 15 and 17, respectively, and a substitution of A to T at position 20. The nucleotide composition of this repeat region is 71% A+T, which is the lowest in the whole mitogenome. Repetitive sequences have been commonly found in the A+T-rich region of mitogenomes (Zhang and Hewitt, 1997), with length variation due to a variable number of copies of repeats. It has been suggested that the origin and persistence of large repeat units within the A+T-rich region are due to replication errors caused by slippage of the replisome and consequent duplication in the region between the origin of replication and the slip (Macey et al., 1998). However, Cameron and Whiting (2007) believe that replication-mediated processes are responsible for the observed pattern of repeats, because of the complicated structure of the repeat units and their overlap adjacent to stem/loop regions. Thus, more data will be informative about how repeat units evolve in the A+T-rich region.

Gene rearrangements

With a few notable exceptions, the gene arrangement in the mitogenome is highly conserved within closely related groups of insects. Among the 37 mitochondrial genes, a few tRNA genes are prone to vary in position, particularly those near the A+T-rich region and in the tRNA gene cluster $tRNA^{Ala}$-$tRNA^{Arg}$-$tRNA^{Asn}$-$tRNA^{Ser\text{-}AGN}$-$tRNA^{Glu}$-$tRNA^{Phe}$. In the *L. striatellus* mitogenome, several PCGs and tRNA genes have translocated or inverted compared with *D. yakuba* (Fig. 1).

The results of the phylogenetic analyses show that *L. striatellus* is well grouped with a related species, *G. distinctissima* (Fig. 3), and this confirms that the mitogenome sequence of *L. striatellus* is from the insect. Possible evolutionary mechanisms for the gene rearrangements in *L. striatellus* are discussed in the following paragraphs.

For translocations of mitochondrial genes, a mechanism of tandem duplication and deletion is plausible. This involves the tandem duplication of gene regions, most widely considered to be the result of slipped-strand mispairing during replication, followed by deletion of one of the duplicated regions (Boore, 2000). Intergenic spacer length variation may have arisen through retention of partial duplications (Mcknight and Shaffer, 1997), or through incomplete multiple deletions of redundant genes (Yamauchi et al., 2003) by the duplication-and-deletion mechanism. For this reason, we speculate that the intergenic spacers distributed in the mitogenome of *L. striatellus* may serve as a guide in deducing the derived gene arrangement. There is a 72-bp stretch of unassignable intervening nucleotides between the *nad4-nad4l* gene cluster and $tRNA^{His}$ in the mitogenome of *L. striatellus*. This region is strongly biased towards A+T (84.9%), but its origin is unknown. Although this intergenic sequence does not correspond to any gene that has possibly undergone a duplication/random loss event, homology may have been lost due to mutation events as a consequence of minimal or no selective pressure on the non-coding nucleotides. Therefore, it is possible that this unassignable sequence represents the degenerating vestiges of genes that have undergone duplication/random-loss events, thus leading to the translocations of the *nad4-nad4l* gene cluster and *nad6* (Fig 4). As for the inversions of two tRNA gene clusters ($tRNA^{Cys}$-$tRNA^{Trp}$ and $tRNA^{Pro}$-$tRNA^{Thr}$), intramitochondrial recombination may be responsible (Dowton and Campbell, 2001). Intramitochondrial recombination specifically involves the breaking and rejoining of DNA double strands. After double-strand breaks in two neighbouring loci (e.g., in the region of $tRNA^{Cys}$-$tRNA^{Trp}$), rejoining of the broken ends can produce three possible products: (i) the major and mini circles previously observed; (ii) an unchanged genome; or (iii) a genome with a short, inverted segment (Dowton and Austin, 1999). These facilitate gene rearrangements and

inversions.

It is possible that other mechanisms may be involved in gene rearrangements. For example, a high rate of endogenous DNA damage, due to a high rate of nucleotide substitution, may result in a high rate of double-strand breaks, which may cause illegitimate recombination (Boore, 2000). Of course, other factors not yet discovered may also affect the rate of gene rearrangement in the mitogenomes of animals. Thus, further studies on gene rearrangements in insects should be informative.

Phylogenetic relationships in Hemiptera

The phylogenetic analyses based on mitogenome sequences indicate that Fulgoromorpha and Sternorrhyncha are sister groups, which is inconsistent with some traditional opinions (von Dohlen and Moran, 1995; Campbell et al., 1995). The topology shows Heteroptera as the sister group to all other Hemiptera; Sternorrhyncha as monophyletic and the sister group to Fulgoromorpha; and Cicadomorpha as the sister group to Fulgoromorpha+Sternorrhyncha. These phylogenetic relationships among hemipteran lineages are similar to those postulated by Hamilton (1981). The phylogenetic analysis of Pentatomomorpha based on PCGs is similar to that of Hua et al. (2008), whose study was based on all 37 genes in the mitogenome. Our results similarly support the hypothesis of (Aradoidea+(Pentatomoidea+(Pyrrhocoroidea+(Lygaeoidea+Coreoidea)))), as both the bootstrap values and posterior probabilities strongly support the corresponding node (Fig. 3). This shows that the PCGs have the same resolving power in phylogenetic analyses as the whole mitogenome. Additionally, there was strong bootstrap support for three major clades in Sternorrhyncha: Psylloidea, Aphidoidea, and Aleyrodoidea.

The gene rearrangements found in *L. striatellus* did not cause any change in the phylogenetic affiliations among hemipterans; a monophyletic Fulgoromorpha was well recovered, with strong bootstrap support in both the ML and BI analyses. *Laodelphax striatellus* always grouped together with *G. distinctissima*. It was a coincidence that we chose to study *L. striatellus*, with gene rearrangements, as perhaps many other delphacid mitogenomes would not have an altered gene order. Further studies of mitogenomes within this group or in close relatives are needed. Similarly to *L. striatellus*, four of six species of whiteflies show variation in gene order. Phylogenetic analyses of Aleyrodoidea using mitochondrial and endosymbiont genes showed that the gene rearrangement in related whitefly species occurred in a common ancestor of these species (Thao et al., 2004). Our phylogenetic analyses based on PCGs indicated that Aleyrodoidea is monophyletic (Fig. 3), though significant nucleotide divergences led to long branch attraction by Aleyrodoidea. Based on analyses of 18s rDNA sequences, von Dohlen and Moran (1995) found a high substitution rate in Aleyrodoidea, whose distances to outgroup taxa were greater than those for other Sternorrhyncha to outgroups. Similarly, the clade of Aleyrodoidea is an unusually long branch in the whole hemipteran phylogenetic topology (von Dohlen and Moran, 1995). However, there seem to be no data that circumvent the artifact of long branch attraction. So, the phylogenetic affiliations of Aleyrodoidea should be tested in future work because of the potential for this unusually long branch to cause misleading results (Felsenstein, 1978).

Hemiptera is the largest nonholometabolan insect assemblage, and many taxa within this order are underrepresented in terms of complete mitogenome sequences. For example, no complete mtDNA sequence for a representative of Coleorrhyncha is presently available in NCBI. To further understand hemipteran phylogeny, sequences from more species within this order are required. As more mitogenome sequences become available, reanalysis of the data may be able to better resolve the phylogenetic relationships within Hemiptera.

ACKNOWLEDGMENTS

This study was supported by grants from the National Natural Science Foundation of China (30530110), a grant (No. O529YX5105) from the Key Laboratory of the Zoological Systematics and Evolution of the Chinese Academy of Sciences and the National Science Fund for Fostering Talents in Basic Research (Special Subjects in Animal Taxonomy, NSFC-J0630964/J0109), both awarded to APL.

REFERENCES

Bae JS, Kim I, Sohn HD, Jin BR (2004) The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. Mol Phylogenet Evol 32: 978–985

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27: 1767–1780

Boore JL (2000) The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals. In "Comparative Genomics, Vol 1" Ed by D Sankoff, J Nadeau, Kluwer Academic Press, Dortrecht, pp 133–147

Boore JL, Brown WM (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr Opin Genetics Dev 8: 668–674

Boore JL, Lavrov DV, Brown WM (1998) Gene translocation links insects and crustaceans. Nature 392: 667–668

Cameron SL, Whiting MF (2007) Mitochondrial genomic comparisons of the subterranean termites from the genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). Genome 50: 188–202

Campbell BC, Steffen-Campbell JD, Sorensen JT, Gill RJ (1995) Paraphyly of Homoptera and Auchenorrhyncha inferred from 18S rDNA nucleotide sequences. Syst Entomol 20: 175–194

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J Mol Evol 22: 252–271

Curole JP, Kocher TD (1999) Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol Evol 14: 394–398

Dotson EM, Beard CB (2001) Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. Insect Mol Biol 10: 205–215

Dowton M, Austin AD (1999) Evolutionary dynamics of a mitochondrial rearrangement 'hotspot' in the Hymenoptera. Mol Biol Evol 16: 298–309

Dowton M, Campbell NJH (2001) Intramitochondrial recombination-is it why some mitochondrial genes sleep around? Trends Ecol Evolut 16: 269–271

Felsenstein J (1978) Cases in which parsimony and compatibility methods will be positively misleading. Syst Zool 27: 401–410

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41: 95–98

Hamilton KGA (1981) Morphology and evolution of the rhynchotan

head (Insecta: Hemiptera, Homoptera). Can Entomol 113: 953–974

Hua J, Li M, Dong P, Cui Y, Xie Q, Bu W (2008) Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). BMC Genomics 9: 610

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755

Hwang UW, Friedrich M, Tautz D, Park CJ, Kim W (2001) Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature 413: 154-157

Junqueira ACM, Lessinger AC, Torres TT, da Silva FR, Vettore AL, Arruda P, Azeredo-Espin AM (2004) The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). Gene 339: 7–15

Kim I, Lee EM, Seol KY, Yun EY, Lee YB, Hwang JS, Jin BR (2006) The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). Insect Mol Biol 15: 217–225

Krzywinski J, Grushko OG, Besansky NK (2006) Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol Phylogenet Evol 39: 417–423

Lewis DL, Farr CL, Farquhar AL, Kaguni LS (1994) Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. Mol Biol Evol 11: 523–538

Lowe TD, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25: 955–964

Macey JR, Schulte JA, Larson A, Papenfuss TJ (1998) Tandem duplication via light-strand synthesis may provide aprecursor for mitochondrial genomic rearrangement. Mol Biol Evol 15: 71–75

McKnight ML, Shaffer HB (1997) Large, rapidly evolving intergenic spacers in the mitochondrial DNA of the salamander family Ambystomatidae (Amphibia: Caudata). Mol Biol Evol 14: 1167–1176

Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F (2003b) Hexapod origins: monophyletic or paraphyletic? Science 299: 1887-1889

O'Brien LB, Wilson SW (1985) Planthopper systematics and external morphology. In "The Leafhoppers and Planthoppers" Ed by LR Nault, JG Rodriguez, Wiley, New York, pp 61–102

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817–818

Roehrdanz RL, Degrugillier ME, Black WC (2002) Novel rearrangements of arthropod mitochondrial DNA detected with long-PCR: applications to arthropod phylogeny and estimation. Mol Biol Evol 19: 841–849

Schultheis AS, Weigt LA, Hendricks AC (2002) Arrangement and structural conservation of the mitochondrial control region of two species of Plecoptera: utility of tandem repeat-containing regions in studies of population genetics and evolutionary history. Insect Mol Biol 11: 605–610

Song N, Liang AP (2009) The complete mitochondrial genome sequence of *Geisha distinctissima* (Hemiptera: Flatidae) and comparison with other hemipteran insects. Acta Biochim Biophys Sin 41: 206-216

Stewart JB, Beckenbach AT (2005) Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumariuss* (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome 48: 46–54

Swofford DL (2002) PAUP* : phylogenetic analysis using parsimony (and other methods), version 4. Sinauer Associates, Sunderland, MA

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599

Thao ML, Baumann L, Baumann P (2004) Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). BMC Evol Biol 4: 25

von Dohlen CD, Moran NA (1995) Molecular phylogeny of the Homoptera: a paraphyletic taxon. J Mol Evol 41: 211–223

Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. Int Rev Cytol 141: 173–216

Yamauchi MM, Miya M, Nishida M (2003) Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (Crustacea: Decapoda: Brachyura). Gene 311: 129–135

Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y (2002) Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. Mol Biol Evol 19: 1385–1389

Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol 25: 99–120

Zhang DX, Szymura JM, Hewitt GM (1995) Evolution and structural conservation of the control region of insect mitochondrial DNA. J Mol Evol 40: 382–391

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406–3415

(Received January 31, 2009 / Accepted August 21, 2009)